Artificial Intelligence II

Bayes’ Nets: Independence

[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at http://ai.berkeley.edu.]

Lecturer (Hannover): Prof. Dr. tech. Wolfgang Nejdl
Probability Recap

- **Conditional probability**
  \[ P(x|y) = \frac{P(x, y)}{P(y)} \]

- **Product rule**
  \[ P(x, y) = P(x|y)P(y) \]

- **Chain rule**
  \[
  P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots \\
  = \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1})
  \]

- **X, Y independent if and only if:** \( \forall x, y : P(x, y) = P(x)P(y) \)

- **X and Y are conditionally independent given Z if and only if:**
  \[
  \forall x, y, z : P(x, y|z) = P(x|z)P(y|z) \\
  \]
  \( X \perp Y | Z \)
A Bayes’ net is an efficient encoding of a probabilistic model of a domain

Questions we can ask:

- Inference: given a fixed BN, what is $P(X \mid e)$?
- Representation: given a BN graph, what kinds of distributions can it encode?
- Modeling: what BN is most appropriate for a given domain?
Bayes’ Net Semantics

- A directed, acyclic graph, one node per random variable
- A conditional probability table (CPT) for each node
  - A collection of distributions over X, one for each combination of parents’ values

\[ P(X|a_1 \ldots a_n) \]

- Bayes’ nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment,
    multiply all:

\[ P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|\text{parents}(X_i)) \]
Example: Alarm Network

\[
P(+b, -e, +a, -j, +m) =
\]
Example: Alarm Network

\[ P(+b, -e, +a, -j, +m) = P(+b)P(-e)P(+a|+b, -e)P(-j|+a)P(+m|+a) = 0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7 \]
Size of a Bayes’ Net

- How big is a joint distribution over \(N\) Boolean variables?
  \[2^N\]

- How big is an \(N\)-node net if nodes have up to \(k\) parents?
  \[O(N \times 2^{k+1})\]

- Both give you the power to calculate
  \[P(X_1, X_2, \ldots X_n)\]

- BNs: Huge space savings!

- Also easier to elicit local CPTs

- Also faster to answer queries (coming)
Bayes’ Nets

- Representation
  - Conditional Independences
  - Probabilistic Inference
  - Learning Bayes’ Nets from Data
Conditional Independence

- X and Y are independent if

\[ \forall x, y \quad P(x, y) = P(x)P(y) \quad \implies \quad X \perp Y \]

- X and Y are conditionally independent given Z

\[ \forall x, y, z \quad P(x, y|z) = P(x|z)P(y|z) \quad \implies \quad X \perp Y|Z \]

- (Conditional) independence is a property of a distribution

- Example: \( \text{Alarm} \perp \text{Fire}|\text{Smoke} \)
Bayes Nets: Assumptions

- Assumptions we are required to make to define the Bayes net when given the graph:

  \[ P(x_i | x_1 \cdots x_{i-1}) = P(x_i | \text{parents}(X_i)) \]

- Beyond above “chain rule = Bayes net” conditional independence assumptions
  - Often additional conditional independences
  - They can be read off the graph

- Important for modeling: understand assumptions made when choosing a Bayes net graph
Example

- Conditional independence assumptions directly from simplifications in chain rule:

- Additional implied conditional independence assumptions?
Independence in a BN

- Important question about a BN:
  - Are two nodes independent given certain evidence?
  - If yes, can prove using algebra (tedious in general)
  - If no, can prove with a counter example
  - Example:

    ![Diagram of X, Y, Z connections]

- Question: are X and Z necessarily independent?
  - Answer: no. Example: low pressure causes rain, which causes traffic.
  - X can influence Z, Z can influence X (via Y)
  - Addendum: they could
D-separation: Outline
D-separation: Outline

- Study independence properties for triples
- Analyze complex cases in terms of member triples
- D-separation: a condition / algorithm for answering such queries
Causal Chains

- This configuration is a “causal chain”

\[ P(x, y, z) = P(x)P(y|x)P(z|y) \]

- Guaranteed X independent of Z?  **No!**

- One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.

- Example:

  - Low pressure causes rain causes traffic, high pressure causes no rain causes no traffic

- In numbers:

  \[
  P( +y \mid +x ) = 1, \quad P( -y \mid -x ) = 1,
  P( +z \mid +y ) = 1, \quad P( -z \mid -y ) = 1
  \]
Causal Chains

- This configuration is a “causal chain”

- Guaranteed $X$ independent of $Z$ given $Y$?

\[
P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)} = P(z|y)
\]

Yes!

- Evidence along the chain “blocks” the influence

$X$: Low pressure          $Y$: Rain                          $Z$: Traffic

$P(x, y, z) = P(x)P(y|x)P(z|y)$
Common Cause

- This configuration is a “common cause”

\[ P(x, y, z) = P(y) P(x|y) P(z|y) \]

- Guaranteed X independent of Z? No!

- One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.

- Example:
  - Project due causes both forums busy and lab full
  - In numbers:
    \[ P( +x | +y ) = 1, P( -x | -y ) = 1, \]
    \[ P( +z | +y ) = 1, P( -z | -y ) = 1 \]
This configuration is a “common cause”

\[ P(x, y, z) = P(y)P(x|y)P(z|y) \]

Guaranteed X and Z independent given Y?

\[
P(z|x, y) = \frac{P(x, y, z)}{P(x, y)}
\]

\[
= \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}
\]

\[
= P(z|y)
\]

Yes!

Observing the cause blocks influence between effects.
Common Effect

- Last configuration: two causes of one effect (v-structures)

- Are X and Y independent?
  - Yes: the ballgame and the rain cause traffic, but they are not correlated
  - Still need to prove they must be (try it!)

- Are X and Y independent given Z?
  - No: seeing traffic puts the rain and the ballgame in competition as explanation.

- This is backwards from the other cases
  - Observing an effect activates influence between possible causes.
The General Case
The General Case

- General question: in a given BN, are two variables independent (given evidence)?
- Solution: analyze the graph
- Any complex example can be broken into repetitions of the three canonical cases
Reachability

- Recipe: shade evidence nodes, look for paths in the resulting graph
- Attempt 1: if two nodes are connected by an undirected path not blocked by a shaded node, they are conditionally independent
- Almost works, but not quite
  - Where does it break?
  - Answer: the v-structure at T doesn’t count as a link in a path unless “active”
**Question:** Are X and Y conditionally independent given evidence variables \{Z\}?
- Yes, if X and Y “d-separated” by Z
- Consider all (undirected) paths from X to Y
- No active paths = independence!

**A path is active if each triple is active:**
- Causal chain $A \rightarrow B \rightarrow C$ where B is unobserved (either direction)
- Common cause $A \leftarrow B \rightarrow C$ where B is unobserved
- Common effect (aka v-structure) $A \rightarrow B \leftarrow C$ where B or one of its descendents is observed

**All it takes to block a path is a single inactive segment**
D-Separation

- Query: $X_i \perp\!\!\!\!\!\!\!\!\!\perp X_j \mid \{X_{k_1}, \ldots, X_{k_n}\}$ ?

- Check all (undirected!) paths between $X_i$ and $X_j$

  - If one or more active, then independence not guaranteed
    $X_i \perp\!\!\!\!\!\!\!\!\!\perp X_j \mid \{X_{k_1}, \ldots, X_{k_n}\}$

  - Otherwise (i.e. if all paths are inactive), then independence is guaranteed
    $X_i \perp\!\!\!\!\!\!\!\!\!\perp X_j \mid \{X_{k_1}, \ldots, X_{k_n}\}$
Example

$R \perp B$

$R \perp B | T$

$R \perp B | T'$

Yes

Diagram:

- Node $R$
- Node $B$
- Node $T$
- Node $T'$

Edges:
- $R$ to $T$
- $B$ to $T$
- $T$ to $T'$
Example

$L \perp T' | T$ \quad Yes

$L \perp B$ \quad Yes

$L \perp B | T$

$L \perp B | T'$

$L \perp B | T, R$ \quad Yes
Example

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I’m sad

- Questions:

  \[ T \perp D \]
  \[ T \perp D | R \quad \text{Yes} \]
  \[ T \perp D | R, S \]
Structure Implications

- Given a Bayes net structure, can run d-separation algorithm to build a complete list of conditional independences that are necessarily true of the form

\[ X_i \perp\!\!\!\perp X_j | \{X_{k_1}, \ldots, X_{k_n}\} \]

- This list determines the set of probability distributions that can be represented
Computing All Independences

Compute **ALL THE INDEPENDENCES**!
Topology Limits Distributions

- Given some graph topology \( G \), only certain joint distributions can be encoded.
- The graph structure guarantees certain (conditional) independences.
- (There might be more independence)
- Adding arcs increases the set of distributions, but has several costs.
- Full conditioning can encode any distribution.

\[
\{X \perp Y, X \perp Z, Y \perp Z, \\
X \perp Z \mid Y, X \perp Y \mid Z, Y \perp Z \mid X\}
\]
Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- D-separation gives precise conditional independence guarantees from graph alone
- A Bayes’ net’s joint distribution may have further (conditional) independence that is not detectable until you inspect its specific distribution
Bayes’ Nets

- Representation
- Conditional Independences
- Probabilistic Inference
  - Enumeration (exact, exponential complexity)
  - Variable elimination (exact, worst-case exponential complexity, often better)
  - Probabilistic inference is NP-complete
  - Sampling (approximate)
- Learning Bayes’ Nets from Data